

## Appendix I: Alignment of NP\_003204 with AAC50763 (SEQ ID NO: 7)

BLASTP 2.2.23+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Reference for compositional score matrix adjustment: Stephen F. Altschul, John C. Wootton, E. Michael Gertz, Richa Agarwala, Aleksandr Morgulis, Alejandro A. Schaffer, and Yi-Kuo Yu (2005) "Protein database searches using compositionally adjusted substitution matrices", FEBS J. 272:5101-5109.

RID: Z8246KTB01S

Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects

11,049,075 sequences; 3,763,785,421 total letters

Query= gi|4507427|ref|NP\_003204.1| TEA domain family member 4 [Homo sapiens]  
Length=434

## ALIGNMENTS

>gb|AAC50763.1| transcription factor RTEF-1 [Homo sapiens]  
Length=434

Score = 907 bits (2345), Expect = 0.0, Method: Compositional matrix adjust.  
Identities = 434/434 (100%), Positives = 434/434 (100%), Gaps = 0/434 (0%)

Query	1	MEGTAGTITSNEWSSPTSPEGSTASGGSQALDKPIDNDGEGVWSPDIEQSFQEALAIYPP	60
		MEGTAGTITSNEWSSPTSPEGSTASGGSQALDKPIDNDGEGVWSPDIEQSFQEALAIYPP	
Sbjct	1	MEGTAGTITSNEWSSPTSPEGSTASGGSQALDKPIDNDGEGVWSPDIEQSFQEALAIYPP	60
Query	61	CGRRKIILSDEGKMYGRNELIARYIKLRTGKTRTRKQVSSHIQVLARRKAREIQAKLKDQ	120
		CGRRKIILSDEGKMYGRNELIARYIKLRTGKTRTRKQVSSHIQVLARRKAREIQAKLKDQ	
Sbjct	61	CGRRKIILSDEGKMYGRNELIARYIKLRTGKTRTRKQVSSHIQVLARRKAREIQAKLKDQ	120
Query	121	AAKDKALQSMAAMSSAQIISATAFHSSMRLARGPGRPAVSGFWQGALPGAETSHDVKPF	180
		AAKDKALQSMAAMSSAQIISATAFHSSMRLARGPGRPAVSGFWQGALPGAETSHDVKPF	
Sbjct	121	AAKDKALQSMAAMSSAQIISATAFHSSMRLARGPGRPAVSGFWQGALPGAETSHDVKPF	180
Query	181	SQQTYAVQPPLPLPGFESPAGPAPSPSAPPAPPWQGRRRGSSKLWMLEFSAFLEQQQDPD	240
		SQQTYAVQPPLPLPGFESPAGPAPSPSAPPAPPWQGRRRGSSKLWMLEFSAFLEQQQDPD	
Sbjct	181	SQQTYAVQPPLPLPGFESPAGPAPSPSAPPAPPWQGRRRGSSKLWMLEFSAFLEQQQDPD	240
Query	241	TYNKHLFVHIGQSSPSYLRPYLEAVDIRQIYDKFPEKKGGLKDLFERGPSNAFFLVKFWA	300
		TYNKHLFVHIGQSSPSYLRPYLEAVDIRQIYDKFPEKKGGLKDLFERGPSNAFFLVKFWA	
Sbjct	241	TYNKHLFVHIGQSSPSYLRPYLEAVDIRQIYDKFPEKKGGLKDLFERGPSNAFFLVKFWA	300
Query	301	DLNTNIEDEGSSFYGVSSQYESPENMIITCSTKVC SFGKQVVEKVET EYARYENGHYSYR	360
		DLNTNIEDEGSSFYGVSSQYESPENMIITCSTKVC SFGKQVVEKVET EYARYENGHYSYR	
Sbjct	301	DLNTNIEDEGSSFYGVSSQYESPENMIITCSTKVC SFGKQVVEKVET EYARYENGHYSYR	360
Query	361	IHRSP LCEYMINFIHKLKHLPEKYMNSVLENFTILQVVTNRDTQETLLCIAYVFEVSAS	420

```
Sbjct  361  IHRSPICEYMINFIHKLKHLPEKYMNSVLENFTILQVVTNRDTQETLLCIAYVFEVSAS  420
Query  421  EHGAQHHiYRLVKE  434
          EHGAQHHiYRLVKE
Sbjct  421  EHGAQHHiYRLVKE  434
```